Estadísticas de usuarios en una biblioteca virtual. El caso de la biblioteca virtual de la red *Clacso*

Por **Dominique Babini**, **Florencia Vergara Rossi** y **Gustavo Archuby**

Babini, Dominique; **Vergara Rossi, Florencia**; **Archuby, Gustavo**. "Estadísticas de usuarios en una biblioteca virtual. El caso de la biblioteca virtual de la red *Clacso*". En: *El profesional de la información*, 2007, enero-febrero, v. 16, n. 1, pp. 57-61.
http://dx.doi.org/10.3145/epi.2007.ene.07

Presentado en:
4ª Jornada sobre la biblioteca digital universitaria–JBDU 2006,
Mendoza, Argentina, 19 y 20 de octubre de 2006

PARTICIPAR EN LA MAYOR BIBLIOTECA que ha tenido hasta ahora la humanidad, la web, es una gran oportunidad para llegar con nuestros servicios a nuevos públicos a medida que se democratice el acceso tanto a los ordenadores como a internet en las comunidades a las cuales queremos alcanzar.

Para una biblioteca esto plantea a su vez grandes desafíos y una buena dosis de espíritu aventurero pues es un espacio nuevo donde los valores, normas y formas de trabajar están en permanente debate y experimentación. Considerando el volumen de información de esta "gran biblioteca", existen buscadores (*Google* y *Yahoo*, entre otros) cuyos robots se dedican a indizar los contenidos de las páginas alojadas en los servidores que las instituciones y personas publican.

Con los servicios tradicionales, el número de usuarios a los que se puede llegar está en relación a su ubicación geográfica, su horario de atención al público, las posibilidades de realizar préstamos interbibliotecarios y de atender a consultas de otras ciudades y países.

En cambio, si ampliamos los servicios con atención en línea, los usuarios a los que se puede llegar tiene relación con la visibilidad de la biblioteca en los buscadores y la disponibilidad de contenidos en texto completo en temas e idiomas de interés entre quienes buscan información en la web.

Son dos universos distintos, con niveles de demanda totalmente diferentes, y es por eso que la biblioteca tiene que definir claramente sus objetivos y el espectro de usuarios a los que se dirige para tenerlos en cuenta en el momento de organizar las actividades y asignar los recursos para tener servicios en línea.

La biblioteca virtual de la *Red Clacso*

En el caso que analizamos aquí, la biblioteca tradicional de *Clacso* (*Consejo Latinoamericano de Ciencias Sociales*) atendía algunas consultas por día en su sede en Buenos Aires, principalmente de usuarios de la misma ciudad. Sin embargo, es un organismo con mandato regional y debe considerar peticiones de todos los países en América Latina y el Caribe, por lo que aprovechar el surgimiento de la web fue un paso lógico para llegar con nuestros servicios a toda la región. Ocho años después de esa decisión los resultados confirman que fue el camino adecuado, pues se reciben más de 100.000 consultas por mes y los principales países que acceden a la información están dentro de América Latina y el Caribe.

http://www.clacso.org.ar/biblioteca

Clacso es una institución internacional no-gubernamental, con relaciones formales de consulta con la *Unesco*. Creada en 1967, hoy agrupa más de 170 centros de investigación y programas de postgrado en ciencias sociales en 21 países de América Latina y el Caribe. La biblioteca virtual surge en 1998 por la necesidad de dar acceso vía web a bases de datos (bibliográficas, de investigaciones y especialistas) y a textos completos (libros, artículos, documentos de trabajo, ponencias) que reflejan la producción de sus centros miembros (hoy la colección tiene más de 7.000 textos digitales). Estos centros producen gran cantidad de publicaciones que ven limitada su difusión internacional debido a las reducidas tiradas, los altos costos del correo y los escasos presupuestos para compra de publicaciones en las bibliotecas. Por este motivo, la mayoría de los centros de investigación de ciencias sociales optaron por la difusión vía web, que es por lo general en acceso abierto (*open access*) con licencias públicas gratuitas de *creative commons* que permiten consultas sin cargo pero exigen citar al autor y su obra, prohibiéndose el uso comercial.

http://www.creativecommons.org

A diferencia de los usuarios que visitan las bibliotecas tradicionales, la mayoría (más del 90%) de los que consultan la web llegan desde los grandes buscadores (*Google*, *Yahoo*, etc.) sin siquiera conocer la biblioteca y sin necesidad de pasar por su *home* o página principal.

http://www.clacso.org.ar/biblioteca

El acceso más fácil e igualitario es señalado como uno de los beneficios de los servicios que las bibliotecas brindan vía web, aun-

que medir e interpretar ese acceso y uso es un proceso complejo que aún no está normalizado internacionalmente, como dice **Peterson Bishop**. En todo caso, son necesarias estrategias de medición para evaluar la transición entre los servicios tradicionales y los servicios a través de internet (**Borgman**; **Larsen,** 2003).

Como también es sabido, desconocemos a la mayoría de los usuarios, pues las consultas las realizan solos, y no tenemos contacto físico o directo con ellos.

En toda biblioteca las estadísticas de uso son básicas para la toma de decisiones. En la virtual, la adquisición de esta información se hace analizando el archivo de logs (generalmente convertido en una base de datos). Existen programas tanto de software libre como comerciales diseñados especialmente para esto.

"Las estadísticas de uso son una herramienta básica para la toma de decisiones"

Estadísticas de usuarios

Es importante conocer los movimientos en el sitio: qué temas solicitan, desde qué país, en qué página decide retirarse de la visita al servicio, etc. Recordemos que la ley protege la privacidad del usuario y no podemos identificarlo por nombre y apellido, a menos que así lo haga él, pero sí es posible conocer su dirección ip lo cual permite generalmente identificar su país de origen y a veces conocer la institución desde donde se realiza la búsqueda (aunque muchas veces queda oculto detrás de un *firewall*). Éstas son estadísticas que señalan tendencias.

Cada vez que alguien accede, la visita se registra con ciertos datos básicos de los movimientos dentro del sitio.

Uso de las estadísticas de una biblioteca virtual

Una vez que tenemos estadísticas del servicio, surgen diversos usos que podemos dar a esta información en apoyo a la gestión interna y externa de la biblioteca virtual, entre otros:

Para la gestión interna es relevante:

—Conocer la cantidad de consultas que recibe la biblioteca por día, mes y año.

—Saber los horarios del día de mayor y menor demanda.

—Establecer el origen geográfico de la consulta, con un listado de países.

—Temas solicitados por los usuarios.

—Textos más pedidos cada mes y cantidad de veces.

—Ranking de las páginas más demandadas por los usuarios.

—Páginas más utilizadas para el ingreso y para el regreso al servicio.

—Cantidad de demandas no satisfechas por problemas con las páginas (error, servicio no disponible momentáneamente, página no encontrada, etc.).

—Kb bajados por día, mes y año.

Para los centros cooperantes que aportan textos completos, así como para el público en general es relevante conocer:

—La cantidad de veces que un texto ha sido solicitado en un mes determinado.

—El número de consultas que recibe la biblioteca por mes.

—Conocer los países que más uso hacen del servicio.

—Temas requeridos por los usuarios.

Asimismo, con diversos objetivos, cada mes se envía una síntesis de estas estadísticas a:

—Autoridades de la institución, para informar.

—Directivos de centros cooperantes: para informar y estimular el envío de textos completos para la biblioteca virtual.

—Personal involucrado en el servicio, con el objetivo de estimularle.

En las dos siguientes secciones de este artículo se describen dos aplicaciones realizadas con software libre para generar estadísticas de uso en la biblioteca virtual de *Clacso*.

Estadísticas generales de una biblioteca virtual. El software libre *Webalizer*

Los servidores web (*Apache*) de *Clacso*, que utilizan como sistema operativo *Linux*, guardan en un archivo *.log* el historial de la actividad que se produce en él. El tráfico se analiza a partir de este archivo que contabiliza los movimientos dentro de la página, por eso posee una gran riqueza de información sobre el comportamiento de los usuarios. Pero la presentación de la información es un archivo de datos sin procesar y conviene hacerlo con un programa que se encargue de organizarlos convirtiéndolos en una síntesis, en muchos casos representándolos gráficamente.

http://www.mrunix.net/webalizer/

En nuestro caso se utiliza *Webalizer* (figura 1). Es una opción libre y de código abierto que funciona en el servidor donde está alojado el sitio. El programa, una vez que ha procesado la información, muestra una página en la que se ve lo que ha ocurrido durante un mes determinado, pudiendo consultar también períodos anteriores. Es un procesador rápido, pero posee limitaciones con respecto a otros interpretadores de datos más poderosos pero que son

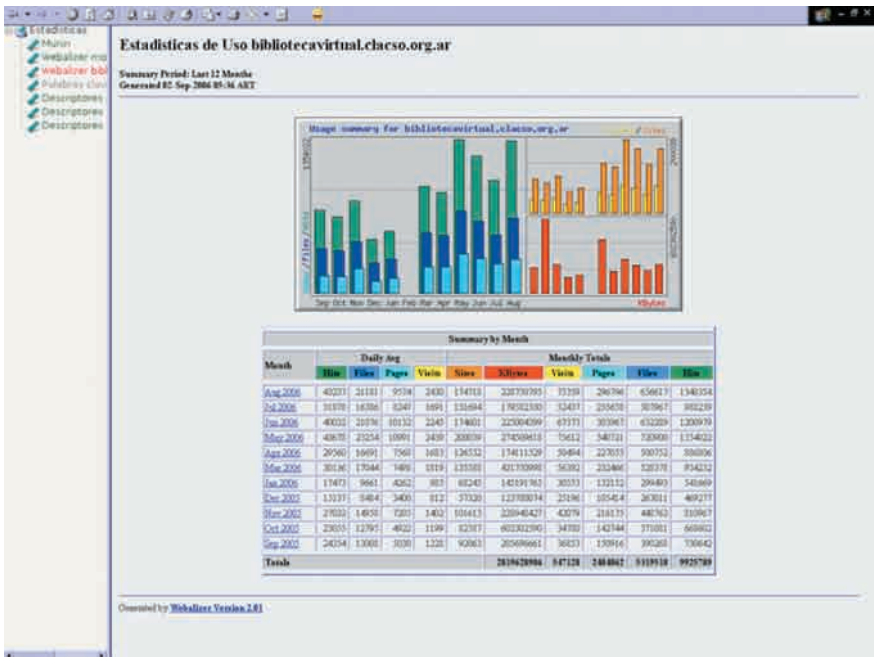

Figura 1: ejemplo de presentación realizada con Webalizer

de pago. Está desarrollado en el lenguaje de programación *C* y se creó para que trabaje sobre plataformas *Linux*; es gratuito y fácil de instalar (para bibliotecas que funcionan con servidores *Windows*, existen otras opciones de programas libres para generar estadísticas).

Podemos obtener y realizar ranking sobre los siguientes aspectos del tráfico: cantidad de visitas por mes, promedio diario, horas de mayor cantidad de tráfico, palabras clave, desde qué páginas acceden los usuarios, países desde donde consultan, cual de las páginas tiene más visitas, cuáles son las últimas que visitan antes de irse; entre otras variables que puede analizar.

Con toda esta información se prepara una página especial, accesible al público desde el sitio con los siguientes datos: la cantidad total de visitantes por mes, el promedio por día y desde qué países se visita más la biblioteca virtual, las 30 palabras claves más buscadas y los 50 textos completos más solicitados del mes.

http://www.clacso.org.ar/biblioteca/Members/estadistica

Este resumen estadístico se actualiza en línea ya que se encuentra diseñado con el programa libre y gratuito de administración de contenidos *Plone* (figura 2).

http://www.plone.org

De los países que más visitan la biblioteca virtual se contabilizan los primeros 30. Para la gestión externa de la institución es importante saber los porcentajes de países de América Latina, y también cómo es el impacto en el resto de las regiones interesadas por los contenidos. Es necesario mencionar que, en general, por las características de internet y de los programas más sencillos de estadísticas, un número importante de consultas quedan sin clasificar por país pues la ip de la máquina que envía la consulta no tiene asignado el país de origen.

Dentro de los servicios que se ofrecen desde esta biblioteca virtual hay tres bases de datos y textos completos; *Webalizer* extrae, consigna y contabiliza las palabras por las que los usuarios realizan consultas dentro de estas plataformas, pudiendo así obtener los temas más buscados. Del resultado total comentamos los primeros 30.

Por el resultado de las estadísticas propias y de aquellas generadas por los distintos buscadores, sabemos que se relevan los contenidos en nuestros servidores periódicamente y están bien posicionados. Esto se debe principalmente a que al encabezado de los archivos en pdf, rtf, doc, se lo modifica, agregándoles la cita bibliográfica en el primer renglón de la página, lo que ayuda a los robots de los buscadores a tomar los datos muy fácilmente y posicionarlos mejor como resultado de búsquedas. Además, incluir la cita bibliográfica en los documentos

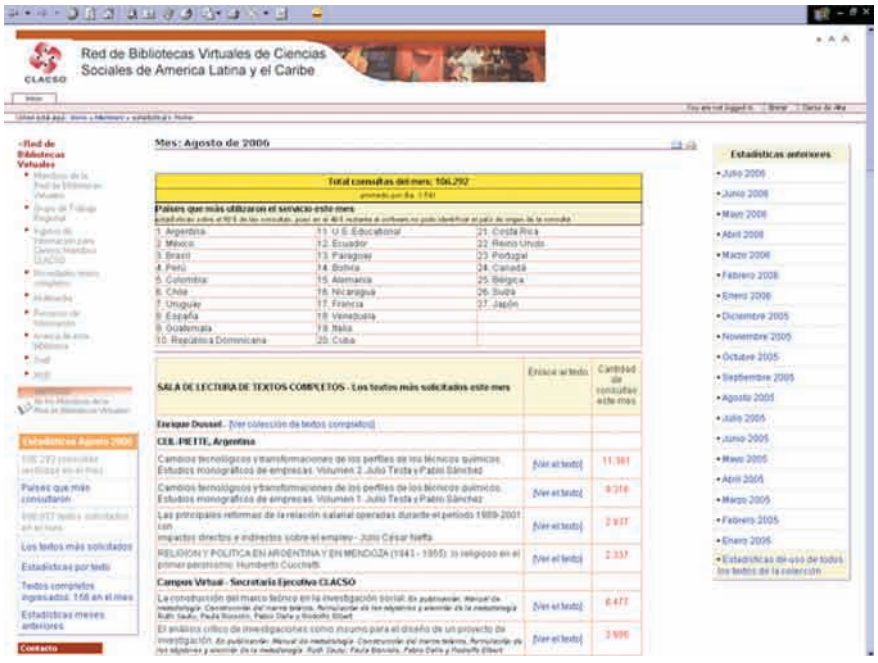

Figura 2: ejemplo de presentación de síntesis estadísticas mensuales

ayuda a los usuarios a citar los trabajos correctamente.

Estadísticas de cada publicación analítica de una biblioteca virtual con software libre

2.1. Estadísticas de textos en formato pdf

En uno de los servidores de la biblioteca virtual, dentro de una misma carpeta, se encuentra el volumen de textos allí alojados, organizados por país, por centro miembro y por programa de la *Secretaría Ejecutiva de Clacso*.

Para tomar datos estadísticos de la cantidad de consultas realizadas sobre los textos digitales se desarrolló un *script* (figura 3) que accede a la información del archivo *.log* y se almacena en una base de datos especialmente creada para guardar y sumar los resultados, consignando por cada texto un total mensual de descargas desde internet. Por otro lado se hizo una interfaz de consulta para el público en general.

El programa para realizar estas estadísticas es un desarrollo propio[1] que se encuentra disponible sin cargo en la sección "Recursos" del sitio web de la biblioteca virtual.

a. Descripción del proceso:

—*Script* que *parsea* (o analiza). Inicialmente se lee cada línea del *.log*, y se obtiene el nombre del archivo al que accedió el usuario.

—Base de datos *MySQL*. Se añade a la base de datos un acceso al documento.

—Interfaz gráfica de consulta. Es un servicio más para los autores que podrán consultar cada mes la cantidad de descargas de sus obras.

De esta base de datos se extraen los 50 documentos más consultados, indicando su título, autor, editor y cantidad de consultas de ese mes, información que también se incluye en las estadísticas mensuales con un enlace al texto.

2.2. Estadísticas de textos digitales analíticos en plataforma *Greenstone*

Cuando la plataforma de una biblioteca virtual permite indizar analíticas (artículo de revista, capítulo de libro, ponencia en congreso, etc.) es natural pensar en el registro de uso de los documentos a este nivel y, en función de éste, realizar un análisis que permita:

"Cuando la plataforma de una biblioteca virtual permite indizar analíticas es natural pensar en tener estadísticas a este nivel"

—En general conocer temas y autores más leídos; en particular, en obras colaborativas capítulos más leídos.

—Tomar la decisión de reeditar una obra o partir de la búsqueda de los usuarios de cierta temática, realizar una investigación sobre la misma, o abrir un concurso para la realización de trabajos en el área o, en el caso de un autor muy popular, solicitar que escriba un artículo en una próxima publicación, etc. Con esta información se puede pensar en profundizar en ciertas temáticas, levantar secciones o cambiar "políticas" para que determinadas secciones poco "leídas" dejen de serlo.

Cabe aclarar que son temas o artículos más leídos y no más buscados, ya que lo que se computa es el acceso al texto completo de un documento y no una palabra colocada en el campo de entrada de un buscador.

Los textos digitales están montados con el programa *Greenstone Digital Library* (*GSDL*) que realiza búsquedas en el texto completo y en metadatos, generando un archivo *.log* que registra los accesos de los usuarios a las partes del documento (analíticas). Este registro es muy rudimentario, pero es la única información que se posee y es a partir de la cual se debe trabajar para obtener la información que buscará estadísticas de uso de la colección a nivel de analíticas. Los archivos *Greenstone* utilizados para las estadísticas son: de log, estructura de directorios y doc.xml.

http://www.gsdl.org

El programa para realizar las estadísticas es también un desarrollo propio y es posible descargarlo gratuitamente.

a. Descripción del proceso

Del archivo de registro de uso no se obtiene exactamente el documento accedido por el usuario, pero sí se identifica el archivo en formato xml que utiliza *Greenstone* para almacenar la información estructuradamente, del cual se puede obte-

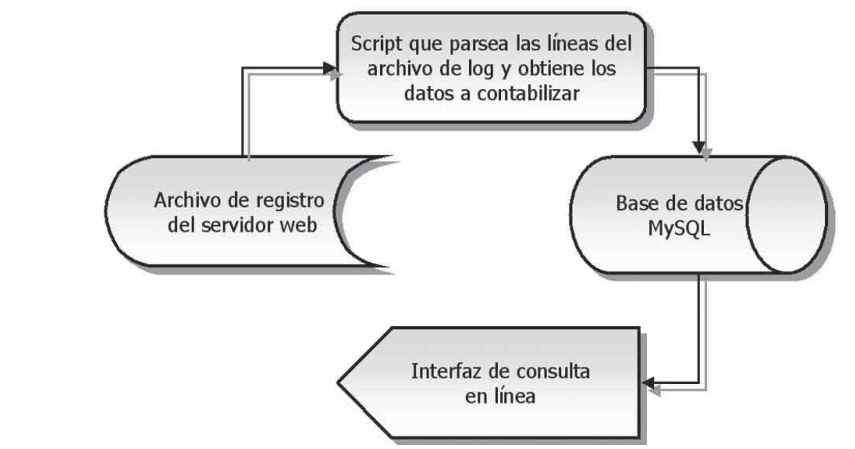

Figura 3: esquema del funcionamiento del script

ner el nombre del documento y de la sección (capítulo), autor, etc. Por tanto, el desarrollo consta a grandes rasgos de 2 partes, un *script* que analiza el *.log*, guarda los resultados en una base de datos y una interfaz de consulta:

—Inicialmente se lee el archivo de registro de uso, se analiza cada línea y se obtiene, por cada una, el nombre del archivo xml y la sección (capítulo) accedida por el usuario.

—Con esta información se accede al archivo xml que contiene los datos que se necesitan (título de la obra, de la parte, autor de ambas) consiguiéndose tanto el título y autor del documento, como el título y autor de la sección.

—Se suma a la base de datos un acceso para dicho libro y parte.

Este *script* funciona automáticamente de forma periódica para actualizar la base de datos. El esquema de la figura 4 muestra a grandes rasgos su funcionamiento.

b. Herramientas utilizadas:

—*PHP*.

—*PHP xml-parser*: desarrollado por **Manuel Lemos**.

`http://www.ManuelLemos.net/`

—*MySQL*.

Conclusiones

El mayor número de visitantes que recibe una biblioteca virtual que ofrece acceso a textos completos vía web se debe principalmente al tráfico que generan buscadores como *Google* y *Yahoo*, más que al conocimiento por parte de esos visitantes de los servicios que ofrece la biblioteca virtual.

Por ello creemos que es necesario conocer el nuevo público que reciben esas bibliotecas. Una forma de hacerlo es generar estadísticas: cantidad de consultas, países, temas más solicitados, textos más bajados... Son todas variables que una vez analizadas tienen un gran valor agregado para tomar decisiones y para la gestión de fondos.

Existen en el mercado programas comerciales para generar estadísticas, pero nosotros promovemos el uso de programas abiertos y libres pues contribuyen al trabajo colaborativo y creativo. Cuando el software disponible no brinda las funciones necesarias hay que explorar otras alternativas, y en este trabajo se describen ejemplos de ello.

Como casi todas las actividades que realizan las bibliotecas virtuales, el análisis estadístico de los usuarios y de los movimientos es también un campo experimental en pleno desarrollo. Lo importante es compartir el camino con otros y crear juntos los mejores indicadores y herramientas que permitan mejorar los servicios que se ofrecen vía web a la comunidad.

Nota

1. Los desarrollos de software libre en la biblioteca virtual de la red *Clacso* están a cargo del analista **Gustavo Archuby** de la *Universidad de La Plata*, Argentina y de **Juan Grigera**.

gustavoa@fahce.unlp.edu.ar
juan@grigera.com.ar

Bibliografía

Borgman, Christine L.; **Larsen, Ronald**. "ECDL 2003 workshop report: digital library evaluation–metrics, testbeds, and processes". En: *D-lib magazine*, n. 9.
http://www.dlib.org/dlib/september03/09inbrief.html#BORGMAN

Jeng, Judy. *Evaluation of digital library: a bibliography*, 2005.
http://www.scils.rutgers.edu/~judyjeng/evaluation.html

Peterson Bishop, Ann. "Measuring access, use and success in digital libraries". En: *The journal of electronic publishing*, 1998, December, v. 4, n. 2.
http://www.press.umich.edu/jep/04-02/bishop.html

***Dominique Babini**, coordinadora de la Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe, Clacso.*
dbabini@campus.clacso.edu.ar

***Florencia Vergara Rossi**, responsable de las plataformas de la Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe.*
fvergara@campus.clacso.edu.ar

***Gustavo Archuby**, consultor informático de la Biblioteca Virtual de Clacso. Analista de computación (Universidad Nacional de La Plata, UNLP, Argentina). Profesor adjunto Cátedra Capacitación en Informática, UNLP.*
gustavoa@fahce.unlp.edu.ar

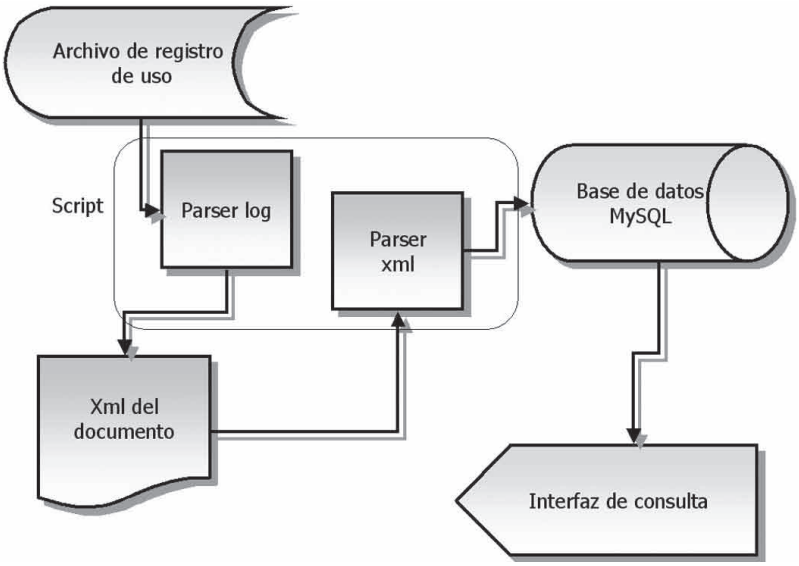

Figura 4